

Robust blind methods using ℓ_p quasi-norms

Abla Kammoun

Abstract—It was shown in a previous work that some blind methods can be made robust to channel order overmodeling by using the ℓ_1 or ℓ_p quasi-norms. However, no theoretical argument has been provided to support this statement. In this work, we study the robustness of subspace blind based methods using ℓ_1 or ℓ_p quasi-norms. For the ℓ_1 norm, we provide the sufficient and necessary condition that the channel should satisfy in order to ensure its identifiability in the noise-less case. We then study its frequency of occurrence, and deduce the effect of channel parameters on the robustness of blind subspace methods using ℓ_1 norms.

I. INTRODUCTION

Despite being proposed several decades ago, blind methods failed so far to be commercially deployed for civil communication products. This can be attributed to several practical difficulties, thereby limiting their use to signal processing applications where training cannot be used. Among the most difficult challenges that need to be addressed is the sensitivity of a number of blind methods to the errors of channel over-modeling, [1], [2]. This is for example the case of conventional subspace methods which are known to exhibit a high sensitivity in case of channel order over-modeling, even in the noiseless case. Actually, in the noiseless case, the channel is identified as the vector that spans the one dimensional kernel of a certain matrix \mathbf{Q} which depends on the second order statistics of the received signal. But, when the channel order is over-estimated, the kernel of matrix \mathbf{Q} is no longer a line but rather a vector space whose dimensions depend on the over-estimated order. The issue that should be raised is thus how to choose the right direction among all the vectors that span the kernel of matrix \mathbf{Q} .

To deal with this problem, a large effort has been investigated in either adding to sensitive methods a new feature that estimates efficiently the channel order [3] or proposing methods that are robust to channel order over-modeling. In this context, a new blind technique for sparse channel estimation has been recently proposed. To select the channel vector, this technique considers the joint utilization of a sparsity criterion which can be measured using a ℓ_p quasi norm ($0 < p \leq 1$). It was noted by simulations that in this way, blind methods like cross relation and deterministic likelihood based techniques become robust to the over-estimation of the channel order [4], [5]. For non-sparse channels, introducing likewise a sparsity criterion shall enhance the channel identifiability probability. The reason is that over-modeling the channel is equivalent to zero-padding the channel vector, which artificially becomes sparse. Moreover, in the case of blind subspace methods, it can be shown that selecting the right direction is equivalent to choosing the vector that maximizes its sparsity.

In this work, we propose to study the robustness of subspace based methods using ℓ_p quasi-norms for non-sparse channels. We derive necessary and sufficient condition for channel identifiability when considering the ℓ_1 norm as well as a sufficient condition if the ℓ_p quasi-norm is used. We then derive a lower bound on the probability that the necessary and sufficient condition holds. This lower bound allows us to study the effect of the system parameters on the channel identifiability probability. For instance, we prove that increasing the number of antennas improves significantly the channel identifiability probability, as opposed to increasing the number of channel coefficients, which tends to reduce it.

We organize this paper as follows: we provide in section II the channel model. Section III is devoted to the derivation of the channel identifiability conditions. Finally, a probabilistic analysis is performed before providing in section V, the simulation results.

II. SYSTEM MODEL AND BLIND SUBSPACE METHODS

Consider a SIMO communication link in which the receiver equipped with M antennas receives data stemming from a single antenna transmitter. The received vector at time k writes as:

$$\mathbf{y}_k = \sum_{l=0}^L \mathbf{h}_l s_{k-l} + \mathbf{v}_k$$

where \mathbf{h}_l is the channel impulse response vector corresponding to the l -th tap and \mathbf{v}_k denotes the additive Gaussian noise vector. Define $\mathbf{h} = [\mathbf{h}_0^T, \dots, \mathbf{h}_L^T]^T$ be the channel vector.

Stacking n observations of vector \mathbf{y} in a $(n+1)M$ vector $\bar{\mathbf{y}} = [\mathbf{y}_k^T, \dots, \mathbf{y}_{k-n}^T]^T$, we will get:

$$\bar{\mathbf{y}} = \mathcal{J}_n(\mathbf{h}) \mathbf{s}_k + \mathbf{v}_k$$

where $\mathcal{J}_n(\mathbf{h})$ is the $M(n+1) \times (L+n+1)$ block-Toeplitz matrix:

$$\mathcal{J}_n(\mathbf{h}) = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_L & \mathbf{0} \\ & \mathbf{h}_0 & \cdots & \mathbf{h}_L \\ & & \ddots & \ddots \\ \mathbf{0} & & \mathbf{h}_0 & \cdots & \mathbf{h}_L \end{bmatrix}$$

Blind methods are based on the second order statistics of the received signal $\bar{\mathbf{y}}_n$. Assuming that:

A.1 The transmitter symbols s_k are independent and identically distributed (i.i.d) with mean zero and variance 1,

A.2 The undergone noise is white with variance σ^2 ,

The covariance matrix of $\bar{\mathbf{y}}_n$ is given by:

$$\mathbf{R} = \mathbb{E} \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^T = \mathcal{J}_n(\mathbf{h}) \mathcal{J}_n(\mathbf{h})^T + \sigma^2 \mathbf{I}_{(n+1)M}.$$

Under the assumption that

A.3 The subchannels of vector \mathbf{h} are real and have no zeros in common. Moreover, $n \geq L$ and $M > 1$,

The rank of $\mathcal{J}_n(\mathbf{h})$ is equal to $L+n+1$. Matrix \mathbf{R} has then exactly $L+n+1$ eigenvalues which correspond to the signal subspace, whereas the remaining eigenvalues correspond to the noise subspace. It admits thus the following singular value decomposition:

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T + \sigma^2\mathbf{N}\mathbf{N}^T$$

where $\mathbf{\Lambda}$ is diagonal with $p = L+n+1$ nonzero diagonal elements and \mathbf{U} and \mathbf{N} are unitary matrices which span respectively the signal and noise spaces. Define $\mathbf{\Pi} = \mathbf{N}\mathbf{N}^T$ the noise projector. In case of the absence of the noise term, the noise projector satisfies:

$$\mathbf{\Pi}\mathcal{J}_n(\mathbf{h}) = \mathbf{0}. \quad (1)$$

Let $\mathbf{\Pi} = [\pi_0, \dots, \pi_M]$. Subspace based methods consists in searching the solution of the following equation:

$$\mathbf{\Pi}\mathcal{J}_N(\mathbf{f}) = \mathbf{0} \quad (2)$$

If the number of coefficients L in \mathbf{f} is accurately known, then there exists up to a constant term only one vector \mathbf{f} of length L satisfying (2). If $L' \geq L$, then, the solutions of (2) belong to a vector space of dimension $L' - L + 1$ spanned by the columns of the following $M(L'+1) \times L' - L + 1$ Toeplitz matrix [1, Lemma2]:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_L & \ddots & \mathbf{h}_0 \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}_L \end{bmatrix}$$

Obviously, the channel vector \mathbf{h} is almost surely the vector that exhibits the highest sparsity among all vectors in the range space of matrix \mathbf{H} . In case the channel order is unknown, the channel vector can be thus recovered by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{f}}{\text{minimize}} && \|\mathbf{f}\|_0 \\ & \text{subject to} && \mathbf{f} = \mathbf{H}\mathbf{s} \\ & && s_1 = 1 \end{aligned} \quad (P_0)$$

where the condition $s_1 = 1$ is introduced to exclude the trivial zero solution (s_1 refers to the first entry of vector \mathbf{s}). Solving (P_0) requires an intractable combinatorial search, thereby reducing its interest for real-time applications. A common used way is to substitute the quasi norm ℓ_0 which is non-convex and non-continuous by the ℓ_p quasi norm, thereby leading to the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{f}}{\text{minimize}} && \|\mathbf{f}\|_p \\ & \text{subject to} && \mathbf{f} = \mathbf{H}\mathbf{s} \\ & && s_1 = 1 \end{aligned} \quad (P_p)$$

where $\|\mathbf{x}\|_p^p = \sum_i |x_i|^p$.

III. CHANNEL IDENTIFIABILITY CONDITIONS

It has been shown by using simulations in [5], that the use of ℓ_1 and ℓ_p norms can enhance the robustness of blind subspace methods. The aim of this paper is to provide theoretical arguments that account for the observed aspect. More explicitly, we will provide necessary and/or sufficient conditions that imply that the solution of problem (P_p) correspond to the zero-padded channel vector in the noiseless case, a case which we refer to as channel identifiability.

Before that, we shall introduce some matrices depending on the channel, which will serve later to formulate the identifiability conditions. Since s_1 has to be set to 1, problem (P_p) could be written as:

$$\underset{\mathbf{g} \in \mathbb{R}^{L'-L}}{\text{minimize}} \quad \|\mathbf{h}_z + \tilde{\mathbf{H}}\mathbf{g}\|_p \quad (P_p)$$

where $\mathbf{h}_z = [\mathbf{h}_1^T, \dots, \mathbf{h}_L^T]^T$ and $\tilde{\mathbf{H}}$ is the $ML' \times L' - L$ block-Toeplitz matrix having the same shape as \mathbf{H} . Partition $\tilde{\mathbf{H}}$ as

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix},$$

where \mathbf{A} is formed by selecting the first ML rows of $\tilde{\mathbf{H}}$ whereas \mathbf{B} correspond to the sub-matrix composed of the last $M(L' - L)$ rows of $\tilde{\mathbf{H}}$, i.e. \mathbf{A} and \mathbf{B} are given by:

$$\mathbf{B} = \begin{bmatrix} \mathbf{h}_L & \mathbf{h}_{L-1} & \cdots & \mathbf{h}_{L-\delta+1} \\ \mathbf{0} & \mathbf{h}_L & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}_L \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{h}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{h}_1 & \ddots & & \\ \vdots & \ddots & & \mathbf{h}_0 \\ \vdots & \vdots & & \vdots \\ \mathbf{h}_{L-1} & \mathbf{h}_{L-2} & \cdots & \mathbf{h}_{L-\delta} \end{bmatrix}.$$

A. Necessary and sufficient condition for ℓ_1 norm

Unlike the ℓ_p quasi-norm ($p < 1$), the ℓ_1 norm is convex. It is thus possible to derive necessary and sufficient condition that implies channel identifiability.

Our theorem is stated as follows:

Theorem 1. Necessary and sufficient condition

Let $\mathbf{v} = [\text{sign}(\mathbf{h}_1)^T, \dots, \text{sign}(\mathbf{h}_L)^T]^T$ and assume that $L > L' - L \geq 1$. Then the necessary and sufficient condition for channel identifiability can be expressed as :

$$\frac{|\mathbf{v}^T \mathbf{A}\mathbf{g}|}{\|\mathbf{B}\mathbf{g}\|_1} \leq 1 \quad \forall \mathbf{g} \in \mathbb{R}^{L'-L} \quad (3)$$

Proof: See Appendix A ■

B. Sufficient condition for ℓ_p quasi-norm

Since the ℓ_p quasi-norm is a non-convex function, the problem might have many local minima. Nevertheless, we still can find a sufficient condition that ensures that the channel can be identified as a local minimum of problem (P_p) . The result is stated in the following theorem:

Theorem 2. Let $\mathbf{v} = p \begin{bmatrix} \text{sign}(\mathbf{h}_1) \\ \vdots \\ \text{sign}(\mathbf{h}_L) \end{bmatrix} \bullet \begin{bmatrix} |\mathbf{h}_1|^{p-1} \\ \vdots \\ |\mathbf{h}_L|^{p-1} \end{bmatrix}$ where \bullet denotes the Hadamard (element by element product). If the following condition is satisfied:

$$\frac{|\mathbf{v}^T \mathbf{A} \mathbf{g}|}{\|\mathbf{B} \mathbf{g}\|_1} \leq 1 \quad \forall \mathbf{g} \in \mathbb{R}^{L'-L} \quad (4)$$

Then, the channel can be identified as a local minimum of (P_p) .

Proof: See Appendix B ■

Remark Note that for each channel realization, there exists p such that the channel identifiability condition (4) occurs.

IV. PROBABILISTIC ANALYSIS

The necessary and sufficient conditions being provided, a natural question that arises is how often they occur. Obviously, this will depend on the distribution of the channel vector elements which have to be chosen. We will consider next the assumption that the channel vector elements are drawn from the Gaussian distribution with zero mean and variance $\frac{1}{L+1}$:

A.4 The entries of the channel vectors $\mathbf{h}_1, \dots, \mathbf{h}_L$ are real Gaussian with zero mean and variance $\frac{1}{L+1}$.

To determine a lower bound on the channel identifiability probability, we will rely on the techniques derived in [6], [7]. Actually, in the same way as in [6], we notice that the necessary and sufficient conditions in the ℓ_1 and ℓ_p norm possesses an equivalent characterization as shown in the following lemma whose proof can be found in [6]:

Lemma 1. Let \mathbf{B} be an $n \times N$ matrix with rank N . For any vector \mathbf{z} define:

$$\|\mathbf{z}\|_B := \sup_{\mathbf{x} \neq 0} \frac{|\mathbf{z}^T \mathbf{x}|}{\|\mathbf{B} \mathbf{x}\|_1}$$

We have then the equivalent characterization :

$$\|\mathbf{z}\|_B = \min \|\mathbf{d}\|_\infty \quad \text{under the constraint} \quad \mathbf{B}^T \mathbf{d} = \mathbf{z}$$

Applying lemma 1, conditions (4) and (3) are equivalent to stating that the solution of the following problem :

$$\min \|\mathbf{d}\|_\infty \quad \text{under the constraint} \quad \mathbf{B}^T \mathbf{d} = \mathbf{A}^T \mathbf{v}$$

achieves a minimum value which is less than 1. In other words, the necessary and sufficient conditions are satisfied if and only if there exists a vector \mathbf{d} with $\|\mathbf{d}\|_\infty \leq 1$ such that $\mathbf{B}^T \mathbf{d} = \mathbf{A}^T \mathbf{v}$.

Since $\text{rank}(\mathbf{B}) = \delta$ almost surely, the channel identifiability will hold if the following conditions are satisfied:

- The image of the cube by matrix \mathbf{B}^T contains a ball of radius α ,

- The vector $\mathbf{A}^T \mathbf{v}$ satisfies $\|\mathbf{A}^T \mathbf{v}\|_2 \leq \alpha$.

Let \mathcal{P} denote the probability that the channel identifiability holds, and E_α^1 and E_α^2 be the events given by:

$$E_\alpha^1 = \{\text{The image of the cube by } \mathbf{B}^T \text{ contains a ball of radius } \alpha\} \\ E_\alpha^2 = \{\|\mathbf{A}^T \mathbf{v}\| \leq \alpha\}$$

Then, \mathcal{P} can be lower-bounded as :

$$\mathcal{P} \geq \mathbb{P} \left\{ \bigcup_{\alpha} E_\alpha^1 \cap E_\alpha^2 \right\} \geq \max_{\alpha} \mathbb{P}(E_\alpha^1 \cap E_\alpha^2).$$

Next, we will consider the computation of the lower bound for the ℓ_1 norm when $\delta = 1$. More explicitly, we show the following theorem:

Theorem 3. For $\delta = 1$, the probability \mathcal{P} that the necessary and sufficient condition (3) occurs is greater than:

$$\mathcal{P} \geq \max_{\varepsilon \in [0,1]} \left(1 - \exp\left(-\frac{M\varepsilon^2}{\pi}\right) \right) \frac{1}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{M(1-\varepsilon)^2}{\pi L}\right) \quad (5)$$

Proof: When $\delta = 1$, it is easy to note that the events E_α^1 and E_α^2 are independent. Therefore,

$$\mathbb{P}(E_\alpha^1 \cap E_\alpha^2) = \mathbb{P}(E_\alpha^1) \mathbb{P}(E_\alpha^2)$$

On the other hand, $\mathbf{v}^T \mathbf{A}$ is a scalar Gaussian random variable with mean 0 and variance $\frac{LM}{L+1}$. Therefore,

$$\mathbb{P}[|\mathbf{v}^T \mathbf{A}| \leq \alpha] = \mathbb{P}\left[|\mathbf{v}^T \mathbf{A}|^2 \leq \alpha^2\right] = \mathbb{P}\left[\frac{|\mathbf{v}^T \mathbf{A}|^2}{LM} \leq \frac{\alpha^2}{LM}\right] \\ = \frac{1}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{\alpha^2(L+1)}{2LM}\right)$$

On the other hand, using standard concentration inequalities (See Theorem 2 in [8]), we have:

$$\mathbb{P}\left[\sqrt{L+1} \|\mathbf{h}_L\|_1 \leq (1-\varepsilon) \sqrt{\frac{2}{\pi}} M\right] \leq \exp\left(-\frac{\varepsilon^2 M}{\pi}\right)$$

Therefore,

$$\mathbb{P}\left[\sqrt{L+1} \|\mathbf{h}_L\|_1 \leq (1-\varepsilon) \sqrt{\frac{2}{\pi}} M\right] \leq \exp\left(-\frac{\varepsilon^2 M}{\pi}\right)$$

As a consequence,

$$\mathbb{P}\left[\|\mathbf{h}_L\|_1 \leq (1-\varepsilon) \sqrt{\frac{2}{\pi(L+1)}} M\right] \leq \exp\left(-\frac{\varepsilon^2 M}{\pi}\right)$$

This gives thus the following probability lower bound :

$$\mathbb{P}\left[\|\mathbf{h}_L\|_1 \geq (1-\varepsilon) \sqrt{\frac{2}{\pi(L+1)}} M\right] \geq 1 - \exp\left(-\frac{M\varepsilon^2}{\pi}\right)$$

Taking $\alpha = (1-\varepsilon) \sqrt{\frac{2}{\pi(L+1)}} M$, we get :

$$\mathcal{P} \geq \max_{\varepsilon \in [0,1]} \left(1 - \exp\left(-\frac{M\varepsilon^2}{\pi}\right) \right) \frac{1}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{M(1-\varepsilon)^2}{\pi L}\right)$$

Interpretation : From the result shown above, one can note that as the number of antennas increases, the probability that the channel identifiability condition occurs tends to 1. On the other side, the order of the channel L might tend to have the opposite effect : it tends to make it fail when it increases.

V. SIMULATION RESULTS

We present here simulation results for the ℓ_1 norm. Fig. 1 displays the effect of the system parameters L and M on the lower bound probability that we have computed by maximizing numerically (5). Fig. 1 shows that increasing the number of antennas tends to enhance the channel identifiability, a result which has been previously mentioned.

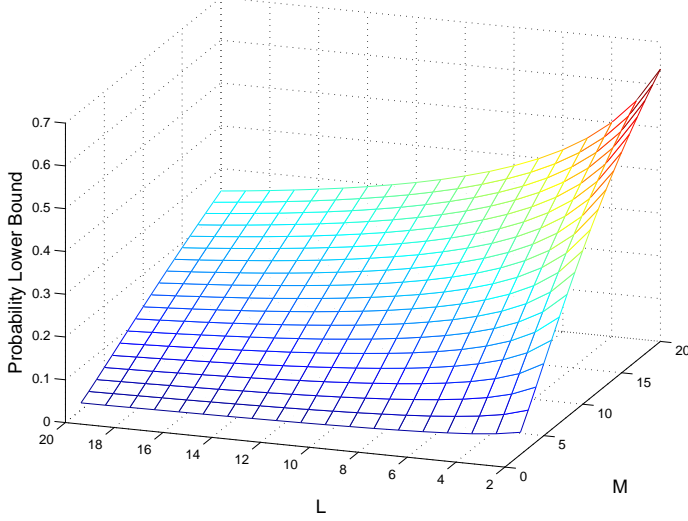


Fig. 1. Impact of the system parameters L and M on the lower bound probability

APPENDIX A PROOF OF THEOREM 1

The proof of theorem 1 will rely on the following mathematical results about the optimization of convex functions, which we provide for sake of completeness.

Definition 1. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a real-valued function. The directional derivative of f at \mathbf{x}_0 is given by :

$$f'(x_0, y) = \inf_{t>0} \frac{f(\mathbf{x}_0 + t\mathbf{y}) - f(\mathbf{x}_0)}{t}$$

Theorem 4. Let f be a convex function defined on a convex set X and $\mathbf{x}_0 \in X$ be a point where f is finite. Then \mathbf{x}_0 is a global minimum point of f if and only of the following condition holds:

$$f'(\mathbf{x}_0, \mathbf{x} - \mathbf{x}_0) \geq 0 \quad \forall \quad \mathbf{x} \in X.$$

We are now in position to establish theorem 1.

Let f be the convex function defined as :

$$f(\mathbf{g}) = \|\mathbf{h}_z + \tilde{\mathbf{H}}\mathbf{g}\|_1.$$

Then, $f(\mathbf{g})$ can be also given by:

$$f(\mathbf{g}) = \|\tilde{\mathbf{h}} + \mathbf{A}\mathbf{g}\|_1 + \|\mathbf{B}\mathbf{g}\|_1$$

where $\tilde{\mathbf{h}} = [\mathbf{h}_1^T, \dots, \mathbf{h}_L^T]^T$.

It is clear that channel identifiability condition occurs if $\mathbf{g} = \mathbf{0}$ is a global minimum. By definition 1, this implies that

$$f'(0, \mathbf{y}) \geq 0 \quad \forall \quad \mathbf{y} \in \mathbb{R}^{L-L} \quad (6)$$

Since

$$\begin{aligned} f'(0, \mathbf{g}) &= \inf_{t>0} \frac{\|\tilde{\mathbf{h}} + t\mathbf{A}\mathbf{g}\|_1 + t\|\mathbf{B}\mathbf{g}\|_1 - \|\tilde{\mathbf{h}}\|_1}{t} \\ &= \mathbf{v}^T \mathbf{A}\mathbf{g} + \|\mathbf{B}\mathbf{g}\|_1. \end{aligned}$$

equation (6) implies that:

$$\frac{|\mathbf{v}^T \mathbf{A}\mathbf{g}|}{\|\mathbf{B}\mathbf{g}\|_1} \leq 1.$$

APPENDIX B PROOF THEOREM 2

To prove theorem 2, it suffices to establish that when condition (4) is satisfied, vector $\mathbf{g} = \mathbf{0}$ is a local minimum of the function

$$f_p(\mathbf{g}) = \|\mathbf{h}_z + \tilde{\mathbf{H}}\mathbf{g}\|_p^p = \|\tilde{\mathbf{h}} + \mathbf{A}\mathbf{g}\|_p^p + \|\mathbf{B}\mathbf{g}\|_p^p$$

where $\tilde{\mathbf{h}} = [\mathbf{h}_1^T, \dots, \mathbf{h}_L^T]^T$. In other words, we need to show that for all $\varepsilon > 0$ there exists a neighborhood in which all vectors \mathbf{g} satisfy $\|\tilde{\mathbf{h}} + \mathbf{A}\mathbf{g}\|_p^p + \|\mathbf{B}\mathbf{g}\|_p^p \geq \|\tilde{\mathbf{h}}\|_p^p$.

Let \tilde{h}_i denote the i -th entry of $\tilde{\mathbf{h}}$ and \mathbf{e}_i be the vector with 1 on the i -th position and 0 elsewhere. Let $\rho_i = \text{sign}(\tilde{h}_i)$. Then, it can be easy to show that :

$$\begin{aligned} |\tilde{h}_i + \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p &= |\tilde{h}_i| + \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p \\ &= |\tilde{h}_i|^p \left| 1 + \frac{\rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g}}{|\tilde{h}_i|} \right|^p \end{aligned}$$

Thereby, the Taylor expansion of $|\tilde{h}_i + \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p$ when \mathbf{g} lies in a neighborhood of $\mathbf{0}$ can be given by :

$$|\tilde{h}_i + \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p = |\tilde{h}_i|^p \left(1 + \frac{p\rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g}}{|\tilde{h}_i|} + \frac{p(p-1)}{2} \frac{|\mathbf{e}_i^T \mathbf{A}\mathbf{g}|^2}{|\tilde{h}_i|^2} \right) + o(\|\mathbf{g}\|^3)$$

Hence,

$$\frac{|\tilde{h}_i + \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p - |\tilde{h}_i|^p - p\rho_i |\tilde{h}_i|^{p-1} \mathbf{e}_i^T \mathbf{A}\mathbf{g}}{p\rho_i |\tilde{h}_i|^{p-1} \mathbf{e}_i^T \mathbf{A}\mathbf{g}} \xrightarrow{\mathbf{g} \rightarrow \mathbf{0}} 0.$$

As a consequence, for all $\varepsilon > 0$, there exists a neighborhood $V_{\varepsilon, i}$ of $\mathbf{0}$ such that :

$$|\tilde{h}_i + \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p \geq |\tilde{h}_i|^p + p|\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g} - \varepsilon p |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{A}\mathbf{g}|$$

We can assume without loss of generality that:

$$\sum_{i=1}^{LM} \rho_i |\tilde{h}_i|^{p-1} \mathbf{e}_i^T \mathbf{A}\mathbf{g} \leq 0.$$

Because, otherwise, one can prove easily for ε small enough, the existence of a neighborhood in which the following desired inequality holds

$$\|\tilde{\mathbf{h}} + \mathbf{A}\mathbf{g}\|_p^p + \|\mathbf{B}\mathbf{g}\|_p^p \geq \|\tilde{\mathbf{h}}\|_p^p.$$

Since $\sum_{i=1}^{LM} p|\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g} = - \left| \sum_{i=1}^{LM} p|\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g} \right|$, we get for all $\mathbf{g} \in \tilde{V}_\varepsilon = \cap_{i=1, \dots, LM} V_{\varepsilon, i}$,

$$\sum_{i=1}^{LM} |\tilde{h}_i + \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g}|^p \geq \sum_{i=1}^{LM} |\tilde{h}_i|^p - \left| \sum_{i=1}^{LM} p|\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{A}\mathbf{g} \right| - \varepsilon p \sum_{i=1}^{LM} |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{A}\mathbf{g}| \quad (7)$$

On the other hand, as $p < 1$, one can easily check that

$$\frac{\|\mathbf{Bg}\|_p^p}{\|\mathbf{Bg}\|_1 - \varepsilon p \sum_{i=1}^{LM} |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{Ag}|} \xrightarrow{\mathbf{g} \rightarrow 0} +\infty$$

Choose ε such that

$$\|\mathbf{Bg}\|_1 \geq \varepsilon p \sum_{i=1}^{LM} |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{Ag}|.$$

This is possible since, with probability equal to one, $\|\mathbf{Bg}\| = 0$ if and only if $\mathbf{g} = 0$. Hence, there exists a neighborhood V_ε such that for all $\mathbf{g} \in V_\varepsilon$, we have:

$$\|\mathbf{Bg}\|_p^p \geq \|\mathbf{Bg}\|_1 - \varepsilon p \sum_{i=1}^{LM} |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{Ag}| \quad (8)$$

On the other hand, the channel identifiability condition (4) implies that

$$\|\mathbf{Bg}\|_1 \geq \left| \sum_{i=1}^{LM} p |\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{Ag} \right| \quad (9)$$

Plugging (8) into (9), we get:

$$\|\mathbf{Bg}\|_p^p \geq \left| \sum_{i=1}^{LM} p |\tilde{h}_i|^{p-1} \rho_i \mathbf{e}_i^T \mathbf{Ag} \right| - \varepsilon p \sum_{i=1}^{LM} |\tilde{h}_i|^{p-1} |\mathbf{e}_i^T \mathbf{Ag}|.$$

This proves by using (7), that for all $\mathbf{g} \in V_\varepsilon \cap \tilde{V}_\varepsilon$, we have:

$$\|\tilde{\mathbf{h}} + \mathbf{Ag}\|_p^p + \|\mathbf{Bg}\|_p^p \geq \|\tilde{\mathbf{h}}\|_p^p,$$

thereby proving the result.

REFERENCES

- [1] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A Subspace Algorithm for Certain Blind Identification Problems," *IEEE Transactions on Information Theory*, vol. 43, no. 2, Mar. 1997.
- [2] L. Perros-Meilhac, E. Moulines, P. Chevalier, and P. Duhamel, "A Parametric Subspace-Based Blind Estimation of a SIMO-FIR With Unknown Channel Order," *SPAWC*, pp. 227–230, 1999.
- [3] A. Gorokhov, M. Kristensson, and B. Ottersten, "Robust Blind Second-Order Deconvolution," *IEEE Signal Processing Letters*, no. 6, pp. 13–16, Jan. 1999.
- [4] A. Aissa-El-Bey and K. Abed-Meraim, "Blind Identification of Sparse SIMO Channels using Maximum A Posteriori Approach," *EUSIPCO*, Aug. 2008.
- [5] A. Aissa-El-Bey and K. Abed-Meraim, "Blind SIMO channel Identification using a Sparsity Criterion," *SPAWC*, 2008.
- [6] R. Gribonval and K. Schnass, "Dictionary Identification : Sparse Matrix Factorization via ℓ_1 Minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3523–3539, 2010.
- [7] R. Gribonval and K. Schnass, "Dictionary Identifiability from Few Training Samples," *EUSIPCO*, Aug. 2008.
- [8] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst, "Average Case Analysis of Multichannel Thresholding," *ICASSP*, 2007.